

The Adoption of Image-Driven Machine Learning for Microstructure Characterization and Materials Design: A Perspective

This pre-print is currently undergoing peer review for publication in JOM.

Arun Baskaran^{1,6} Elizabeth J. Kautz² Aritra Chowdhary³
Wufei Ma⁴ Bulent Yener⁵ Daniel J. Lewis¹

¹Department of Materials Sci. & Engg., Rensselaer Polytechnic Institute, NY, USA

²Pacific Northwest National Laboratory, WA, USA

³Artificial Intelligence, GE Research, NY, USA

⁴Department of Computer Science, Purdue University, IN, USA

⁵Department of Computer Science, Rensselaer Polytechnic Institute, NY, USA

⁶Currently at the Center for Nanoscale Materials, Argonne National Laboratory, IL, USA

Abstract

The recent surge in the adoption of machine learning techniques for materials design, discovery, and characterization has resulted in an increased interest and application of Image Driven Machine Learning (IDML) approaches. In this work, we review the application of IDML to the field of materials characterization. A hierarchy of six action steps is defined which compartmentalizes a problem statement into well-defined modules. The studies reviewed in this work are analyzed through the decisions adopted by them at each of these steps. Such a review permits a granular assessment of the field, for example the impact of IDML on materials characterization at the nanoscale, the number of images in a typical dataset required to train a semantic segmentation model on electron microscopy images, the prevalence of transfer learning in the domain, etc. Finally, we discuss the importance of interpretability and explainability, and provide an overview of two emerging techniques in the field: semantic segmentation and generative adversarial networks.

Keywords: IDML, Microstructure Characterization, Machine Learning, Machine Learning in Material Science

1 Introduction

Microstructure is a broad term that encompasses the description of a material's structure at spatial length scales ranging from millimeters to nanometers. A material's property and performance is strongly influenced by its microstructure, which is often studied by capturing and analyzing images using various microscopy methods. Such micrographs are rich in information about the material's origin and processing history and its chemical make-up. Hence, microstructure visualization, analysis, and interpretation has become ubiquitous in materials science research. Significant advances in both experimental and computational methods have enabled advances in how we visualize and describe material microstructures. For example, three dimensional microstructures can be visualized with techniques such as serial sectioning and atom probe tomography [1, 2], and in situ methods have been developed for direct visualization of atomic-scale phenomena, including oxide growth mechanisms and element redistribution [3, 4]. Computational methods have been at the forefront of many materials research studies due to the need for accelerated materials discovery, design, and development emphasized in the Materials Genome Initiative [5]. Methods including ab initio calculations, density functional theory (DFT), n-point statistics, and machine learning have recently gained more widespread use, enabled by modern (often, open source) algorithms and high-performance computing. More recently, the application of artificial intelligence (AI) has grown significantly in engineering and science domains, particularly in the materials science area. This boom in AI has been referred to as the fourth paradigm of science [6] and the fourth indus-

trial revolution [7], and has enormous potential in altering how materials scientists discover new material systems, predict properties, and even interpret and analyze micrographs.

The topic area of machine learning in materials science, or material informatics, is broad. Reviews of the overall field can be found in [8,9]. Within this broad field, a specific area of importance is the characterization and interpretation of image data using machine learning methods, referred to here as image driven machine learning (IDML). Image data is ubiquitous in materials characterization efforts, thus, there is a strong need to develop models and approaches that can accurately, effectively, and reliably link image data to other parameters of interest (i.e. processing parameters, properties, etc.). In the specific area of microstructure image analysis, machine learning is proving highly valuable for characterizing diverse microstructure data sets paramount to the materials design and discovery process [10,11]. Yet, there is still much to be learned about how to apply machine learning methods to the important tasks of microstructure recognition, characterization, development of processing-microstructure-property relationships, and microstructure design. The motivation of the current review is to provide a perspective on the different aspects involved in the application of IDML techniques for material characterization, by looking at the existing literature. The text is likely to provide a high level overview of the field to material scientists who are interested in the field of applied machine learning, as well as enable the experienced practitioners to obtain a summary of the field or an insight into an allied field.

In this article, we begin by outlining the current state-of-the-art of IDML in materials science. The scope of this review includes image datasets obtained from camera (imaging sample surface at the resolution of naked eye), optical microscopy, electron microscopy, spectroscopy, diffraction patterns, and from simulations of material structure at different length scales. Following this, we define a canonical hierarchy of procedural stages which is used to modularize the IDML studies reviewed in this work. We briefly discuss the importance of interpretability and explainability, towards a widespread adoption of IDML in materials science. Finally, we briefly discuss two emerging techniques that have gained importance in the last couple of years: generative adversarial network and semantic segmentation.

2 Overview of the field

Although it is common for materials scientists to relate microstructure and properties, the availability of microstructural information is a relatively recent discovery compared to the length of time materials technology has been known. C. S. Smith [12] has written extensively on the subject of the discovery of microstructure in materials and cites what is considered to be the first observations of metallic crystallinity through the use of chemical etchants in meteorites by Thompson in 1804 and in Damascus steel by Breant in 1821. The need to quantify the description of these structures was immediately apparent and we refer the reader to C. S. Smith's writings on the history of technology and structure of metals. A significant contribution of Smith's was in the description of polycrystalline structures and an early application of a computational image processing technique to a microscopy image was the use of the intercept method [13] to calculate the average grain size of a polycrystal. Since then, the field has experienced the use of progressively advanced techniques to extract complex insights from microscopy images. While the next section will review journal articles in this field on the basis of their technical aspects, this section will focus on the materials science problems that have been addressed so far using IDML.

In the last decade, IDML has augmented the human expert's understanding of material characterization at a range of length scales spanning multiple orders of magnitudes. At the lower magnification levels, IDML techniques have been applied to image datasets with a field of view in the range of 10-100 μm . Examples of such studies include porous microstructure reconstruction using generative methods [14], in-situ defect detection in melt pools for additive manufacturing [15,16], and segmentation of images generated from x-ray computed tomography and serial sectioning [17]. At comparatively higher magnification levels, IDML has been used to analyze optical and electron microscopy image datasets with a field of view in the range of 10-500 μm . A few examples of such studies include high level classification of images into an appropriate microstructure class ([10,11]), semantic segmentation of images generated from x-ray computed tomography ([18,19]), semantic segmentation of images generated from Scanning Electron Microscopy ([20,21]), quantitative metallographic analysis of microscopy images ([22,23]). At the high end of magnification, IDML has been used to analyze images with a field of view in the range of 10-500 nm. Examples of such studies include high-level classification of chirality of carbon nanotubes from images generated by High-Resolution Transmission Electron Microscopy [24], semantic segmentation of crystallographic defects from images generated by Scanning Transmission Electron Microscopy ([25,26]), nanopar-

ticle segmentation from images generated by liquid-phase Transmission Electron Microscopy [27], and analyzing the grain boundary character [28]. A visualization of characteristic images spanning the length scales discussed above has been shown in Figure 1.

A different perspective about the field can be obtained by extracting the domain specific problem statements from the studies. The most extensive application of IDML in the domain has been towards phase segmentation and quantitative analysis of microstructures. The objectives of a segmentation study include the isolation of specific phases [19,21] and the extraction of quantitative information, such as volume fraction, about a specific phase [23,29]. In [132], a supervised approach for binary edge detection allowed for segmentation of phases with different morphologies than were present in the training dataset. Another common application of IDML in this field is the assignment to an image a feature label that best describes the image, from a pre-determined candidate set of features exhibited by the material. The label can indicate a morphology present in a microstructure image ([10,11,30]), the chirality exhibited by the sample [24], processing history of a given microstructure [31], or crystal structure identification from a Transmission Electron Microscopy image [32]. Assigning a class label can providing the human expert with a high-level overview of a microstructure or be an integral part of a multi-stage machine learning pipeline. IDML has also been used for the purpose of defect detection at different length scales and quality control. Examples of such studies include in-situ defect detection in additive manufacturing [15,16], fracture surface analysis from electron microscopy images [20], and segmentation of crystallographic defects from electron microscopy images [25,26].

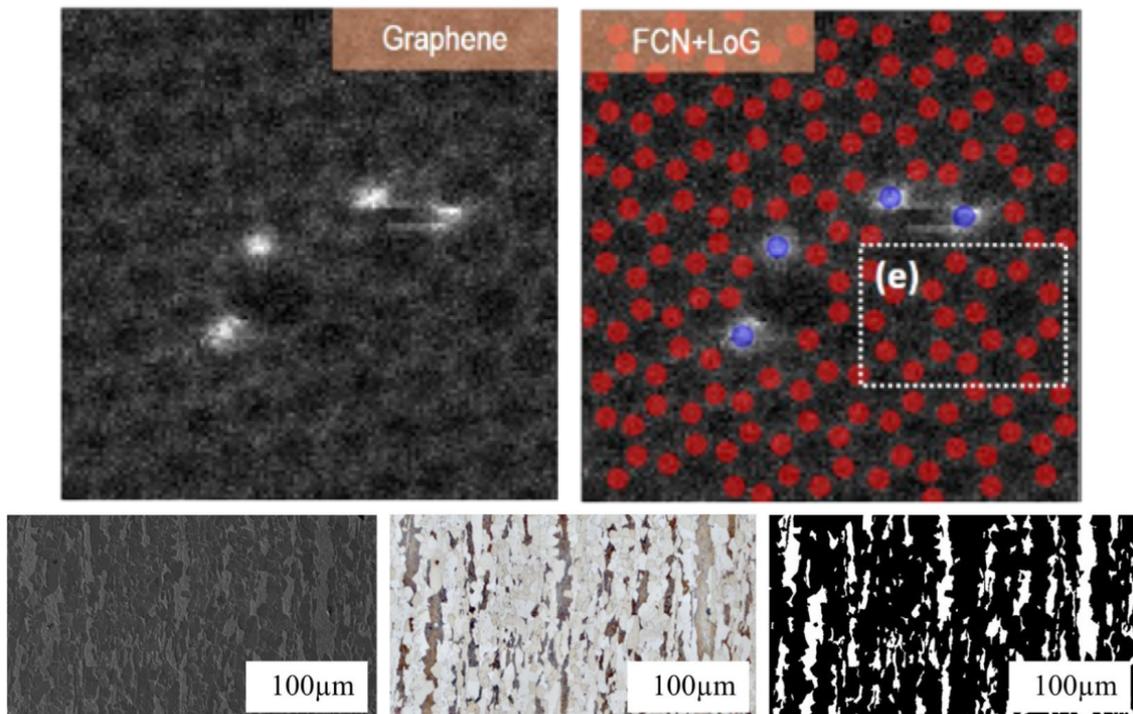


Figure 1: IDML has influenced material analysis across a range of magnifications. (Top) On the smaller end of the spatial length scale, studies have targeted microscopy images with a field of view spanning between 10-500 nm. Figure reproduced with permission from [25]; Copyright American Chemical Society (2017). (Bottom) On the larger end of the spatial length scale, studies have targeted microscopy images with a field of view of 10 μm or more. Figure reproduced with permission from [21]; Creative Commons CC BY.

2.1 Related Work

Machine learning, and more specifically the IDML framework, is considered here to be a subset of the larger domain of Materials Informatics. It is important to note that the terms data-driven, machine learning, and materials informatics are typically used interchangeably. Materials informatics, however, is a large field that encompasses machine learning and statistical models for the quantitative description of microstructures. Several prior works have implemented statistical characterization and reconstruction of microstructures [33–37]. Another important component of

microstructure quantification is the uncertainty quantification, specifically the processing-induced variability in features and its subsequent effect on the material properties. Examples of studies on this topic include [38,39]. While these studies are incredibly valuable and insightful, the application of ML methods for microstructure quantification offer some unique advantages over these statistical techniques, including the ability to incorporate large amounts of data about the material microstructure and analysis of never-before seen data.

Machine learning is now being increasingly used in the materials science domain for a wide range of applications, beyond analysis/classification of images. Although IDML is the focus of this article, it is important to mention that image classification and analysis is merely a subset of the broader field of machine learning in materials science. Hence, we briefly summarize related work on machine learning methods (e.g., neural networks, Support Vector Machines, k-means clustering, random forests, generative networks, and more) applied to several diverse challenges in molecular and materials science fields. In particular, active research areas for ML in materials science include (but are not limited to): accelerated materials design and property prediction [40–46], process optimization [47,48], discovery of structure-property relationships [49,50], construction of potential energy surfaces for molecular dynamics simulations [51–53], prediction of atomic scale properties [54], text mining for knowledge extraction [55], microstructure and materials characterization [10, 11, 25, 30, 56–58], and generation of synthetic microstructure images [31]. Such applications span multiple length scales and a variety of material systems (metals and alloys, oxides, polymers) [59].

The merger of machine learning with materials science is a relatively new and growing field that contributes to the evolution from traditional methodologies, where experimental characterization techniques were used to understand processing-structure-property relationships, to one that is data-driven. This paradigm shift can, in many cases, help accelerate materials science research through a more autonomous, objective, and reliable design and characterization process, which is impacted less by researcher bias and chance discovery [60].

The ability to analyze smaller data sets is important in materials science where a limited number of micrographs are usually available for any single study. Recent advances have led to developments that allow human-level performance in one-shot, or few-shot learning problems [60,61]. Typically, data analysis via deep learning methods require large amounts of data (such as the large image data available through the ImageNet database [62,63]). Although in many materials science studies, researchers are limited to a few data points, an advantage of many micrographs is that several microstructural features of interest are available for interest in a single image (depending on magnification). This one-shot or few-shot learning concept has significant implications for future materials science studies, for example characterization of neutron irradiation or corrosion effects, where limited material is available for analysis due to long lead-time experimentation. In other cases, there exists data from previous studies that may be very limited, or not well understood, for which advanced data analysis methods could be applied, as was done in prior work [46,64].

This relatively recent surge in research efforts at the intersection between materials science and machine learning domains have been enabled by the ever decreasing cost of computing resources. In addition, several frameworks such as Keras [65], PyTorch [66], and TensorFlow [67] have helped materials scientists (amongst other domain scientists such as nuclear engineers, chemists, physicists, etc.) and data scientists readily apply sophisticated machine learning algorithms with relatively low/minimal developmental efforts. The low cost associated with application of machine learning algorithms to various domain-specific research challenges, and the aforementioned benefits enable studies such as those discussed in this article.

3 Component-based breakdown of IDML research efforts

This section provides a modularized approach to discuss the published literature in the field of IDML for materials characterization. A series of six canonical action steps is defined in Figure 2. The review is divided into six components, one for each of the steps outlined in the figure. The granularity of the modularization is chosen so as to keep the individual modules comparable to application of ML in allied fields.

Each action step in the series is approached sequentially, from the bottom-up, in the logical order required to successfully implement a ML technique to the domain. The first step in the series is the definition of a domain-specific goal, i.e., the problem specific to materials science that is being answered by the adoption of IDML. An appropriate problem definition is important to assess the degree of success of the implementation. A review of the field from the perspective of this step is likely to help a beginner to the field with identifying the different components that together make up a rigorous problem definition. well as an experienced practitioner. For an experienced

practitioner, this component of the review provides a summary of the domain-specific problems that have been approached through IDML methods.

The second step in the series is dataset acquisition. Data collection, augmentation, and pre-processing are important steps for a project implementing both computer vision and machine learning techniques. In addition, the allocation of data into training, validation, and testing have also been discussed where appropriate.

Following this, an overview is provided of the classification models and image processing algorithms used among the published literature in the field. Specific focus is provided to the rationale behind model selection that is appropriate for a particular domain-specific goal and dataset. Specific model-based tuning to address domain-specific requirements are highlighted in this subsection.

The fourth component of the review is the training and optimization methods employed by the published studies in the domain. Here as well, the specific modifications to the training process to address domain-specific constraints are highlighted.

The fifth component of the review is the evaluation metrics adopting in the field to assess the performance of the IDML techniques. Evaluation metrics are an assessment of the extent to which the technique addresses a given materials characterization analysis, and hence a greater focus is given to this section. This section is aimed at giving the reader an overview of the benchmark results currently in the field, as well as providing an introduction to a set of metrics that are commonly used in the field.

The final component of the review pertains to a discussion of ways by which research efforts have integrated IDML into their characterization workflows.

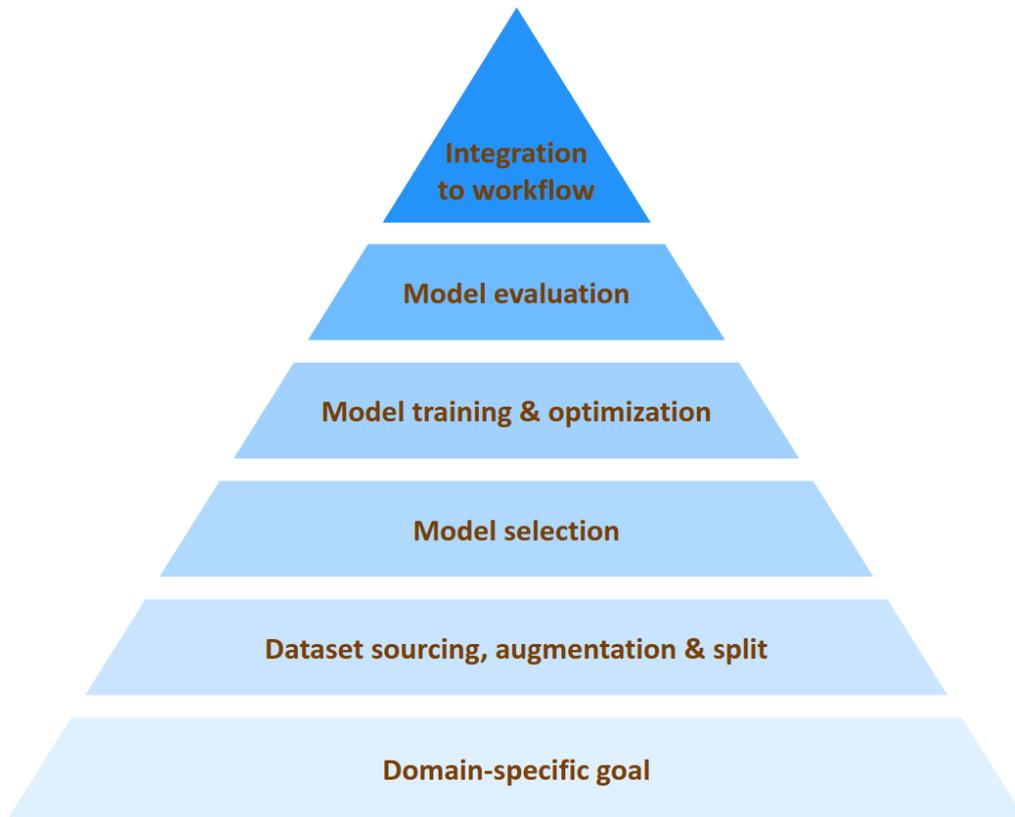


Figure 2: The canonical series of procedural stages used to review the studies in the field of IDML for material characterization. Each stage can be considered as being independent of the others, but crucial to the final performance of the model or algorithm that is being implemented.

3.1 Domain-specific goals

Table 1 outlines the functional goals outlined in the 30 papers reviewed in this study.

One of the main goals of material scientists in relation to the use of computer vision techniques has been towards microstructure classification. The research efforts in this area can be divided into two broad categories, coarse classification and dense classification, depending on the number of

Table 1: Domain-specific goals adopted by a subset of studies that have been reviewed in this work.

Paper	Functional goals
DeCost2015 [10]	Classify microstructure images into one of seven image classes, and develop a bag of visual features
Chowdhury2016 [11]	Binary classification of microstructural features (dendrites) and feature orientations
Zhang2021 [14]	Generate isotropic and anisotropic synthetic 3D porous structures using 2D slices as input
Gobert2018 [15]	In situ defect detection using supervised machine learning, for powder bed fusion (PBF) additive manufacturing
Scime2019 [16]	In-situ defect detection in additive manufacturing melt pools
Stan2020 [17]	Semantic segmentation of computed tomography and serial sectioning images
Strohmann2019 [18]	Semantic segmentation of 3D microstructure of Al-Si
Evsevlev2020 [19]	Deep-learning based semantic segmentation of individual phases from synchrotron x-ray computed tomography images
Tsopanidis2019 [20]	Semantic segmentation of fracture images of MgAl ₂ O ₄ , using a deep-cnn
Azimi2018 [21]	Pixel-wise segmentation of steel microstructure datasets
Campbell2018 [22]	Automated extraction of quantitative data from material microstructures using advanced image processing technique
Agbozo2019 [23]	Quantitative metallographic analysis through object segmentation of SEM images
Forster2020 [24]	Classification of HRTEM images of Carbon Nanotubes into its appropriate chirality
Ziatdinov2017 [25]	Semantic segmentation of defects and subsequent defect structure identification from atomic scale STEM images
Roberts2019 [26]	Semantic segmentation of crystallographic defects from electron micrographs
Yao2020 [27]	Nanoparticle segmentation from liquid-phase TEM images by U-Net
Chan2020 [29]	Automated quantitative analysis of microstructure using unsupervised algorithms
Baskaran2020 [30]	Contextual segmentation of morphological features in titanium alloys using a two-stage machine learning pipeline
Aguiar2019 [32]	Classification of TEM data, atomic resolution images and diffraction data into crystal structures at family level and genera level
Zhu2017 [69]	Particle recognition from cryo-EM datasets
Chen2020 [70]	Instance semantic segmentation of Al alloy metallographic images
DeCost2018 [71]	Semantic segmentation of ultrahigh carbon steel microstructures through deep learning techniques
Furat2019 [72]	Semantic segmentation of computed tomography data of Al-Cu specimens
Vuola2019 [73]	Nuclei segmentation from microscopy images for biomedical imaging using an ensemble model
Hwang2020 [74]	Semantic segmentation of multi-phase composite microstructures
Ma2018 [75]	Semantic segmentation of Al-La microstructure images
Decost2017 [77]	Classify AM feedstock powder images into their respective material system
Yang2019 [85]	Prediction of the microscale elastic strain field in a 3D voxel-based microstructure
Cang2016 [86]	Microstructure reconstruction from a low-dimensional representation
Arganda-Carreras2017 [87]	Develop a library for integration of machine learning methods available in WEKA into image processing toolkit Fiji for biological and non-biological specimens.
Ciresan2012 [88]	Binary membrane segmentation of neuronal structures in stacks of electron microscopy images
Wang2013 [89]	Binary membrane segmentation of neuronal structures in stacks of electron microscopy images
Haan 2019 [90]	Super-resolving low-resolution SEM images through a generative adversarial network
Kaufmann2020 [91]	Crystal structure identification from EBSD diffraction patterns using 6 deep learning
Madsen2018 [92]	Semantic segmentation of HRTEM images for local structure recognition

pixels per classification label. The first category includes classification models that assign one label per image in the dataset. The label that is chosen is the one that best represents the microstructure in the image, such as “dendritic”, “lamellar”, etc. Examples of previous studies that perform such a classification include [10], [11], [21], [30], and [68]. Dense classification models typically assign one label per pixel in a given image. In this case, the label describes the phase of the microstructure at the spatial location represented by the pixel. Examples of research articles that perform this classification include [17–20, 22, 23, 29, 69–76]. A fast emerging technique in this category is semantic segmentation, which can be summarized as pixel classification performed in the context of the whole image. It is noted that there also exists models that perform classification at a scale that is in between the two extremes stated above, such as the prediction of bounding boxes using a Region Proposal Network (RPN) in [70]. Figure 3 shows characteristic examples of application of the different types of classification discussed above to characterization analysis.

These styles of classification have been used by the domain experts towards two broad objectives, object/class detection and extraction of quantitative information from characterization results. The object (for example, precipitate) or class (for example, dendritic) detection is typically performed by either coarse-grained classification or semantic segmentation. Another use for object detection has been the detection of defects and abnormalities. The goal for such a study would be binary classification into “defect” and “no defect”. The classification result is typically integrated with the feedback for process-correction or quality-control. Examples of research articles with this goal include [15, 16, 26, 77–80]. Among the research articles cited above as examples of pixel-level classification, one of the goals is to aggregate the classification into developing a quantitative finger-print of the microstructure, such as area fraction of a particular phase or the number of precipitates in the field of view represented by the image. Both object detection and quantitative analysis of the microstructure potentially aid the study of a process-structure relationship for the material.

Among the studies that have performed quantitative microstructure analysis, a subset of them have developed a lower-dimensional representation for microstructures. Techniques like Principal Component Analysis are used to identify the most discriminative features for the dataset. Such a representation could be an end-goal in itself, like in [68, 81–84, 86, 93], or could be developed in order to improve the classification performance, like in [11, 31].

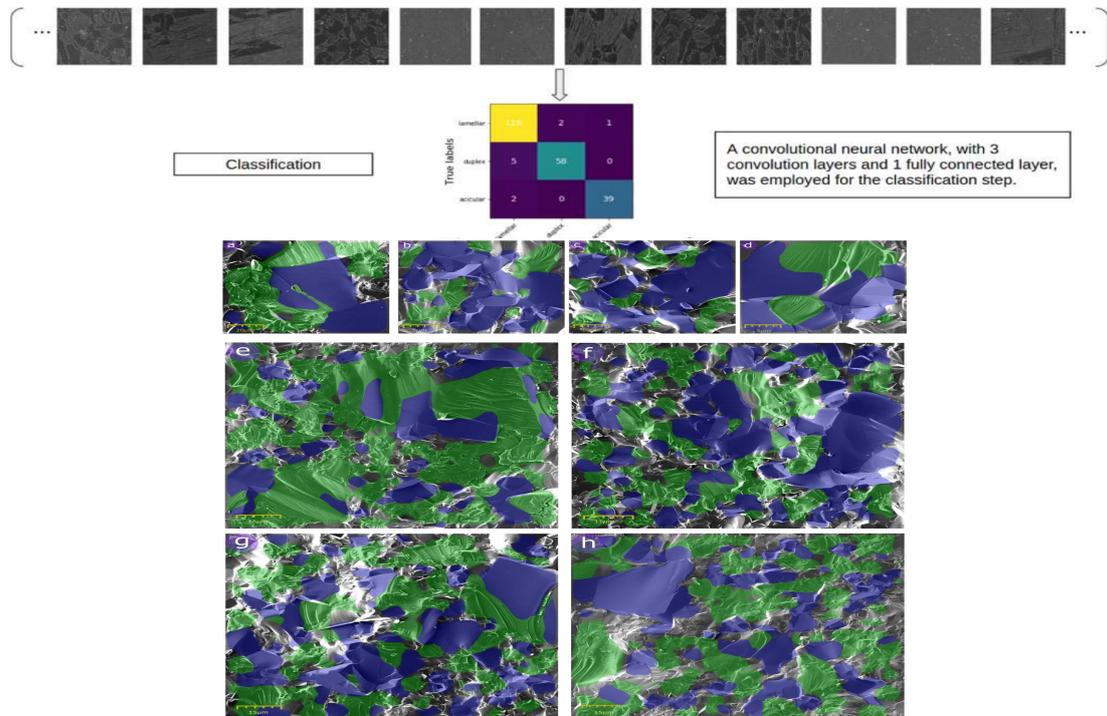


Figure 3: Distinction in the types of classification that can be performed, based on the number of pixels per classification label. (Top) Coarse classification, in which a label is assigned to each image in the dataset. Figure reproduced with permission from [30]; Copyright Elsevier (2020). The example image is extracted from [30]. (Bottom) Dense classification (or pixel-based classification), in which a label is assigned to each pixel in the image. Figure reproduced with permission from [20]; Copyright Elsevier (2020).

3.2 Datasets

The dataset characteristics, such as its size and distribution, have a strong influence on the overall performance of the model and/or algorithm. Most applications of IDML in the materials science domain (not including the studies involving microstructure reconstruction) were performed on real data, i.e. data generated from real material samples. A few exceptions to this can be found in studies such as [24, 25, 31, 32, 85], where simulated data either comprised the entire training dataset or supplemented the data obtained from a real sample. The real data were generated from a wide variety of sources, which have been briefly summarized in Section 2. One can observe a significant variation in dataset sizes among the published literature in this field. There is no rigid rule to determine the appropriate size of the training dataset. It depends on factors such as the problem statement, variation in the dataset, choice of model, required level of performance, etc. Naturally, coarse classification and dense classification studies require different number of images for training. Since the former has a greater number of pixels per classification label than the latter, it requires a comparatively larger dataset. Even among studies that have employed a coarse classification model, a significant variation in dataset size can be observed. For example, while [10] uses 105 images to train a SVM for classification into 7 classes, [11] employs 528 images for a binary classification task using different models, and [30] uses 1225 images to train a CNN for 3-class classification. To the best of our knowledge, the largest dataset used in the domain of materials characterization has $1.3e+6$ images and is reported in [24]. Similarly, a variation in dataset sizes can also be observed among studies that have used dense classification models. For example, while the reported datasets in [23, 70, 71] are 50, 100, and 24 respectively, the reported datasets in [21, 25] are 5093 and 2000, respectively.

In addition to the size of the overall dataset, it is also important to appropriately split the dataset for the purposes of training, validation, and testing. A good evaluation of the performance of a trained model is its performance on data that was not used for training. A general trend observed in the studies reviewed in this work is that the data split depends on the size of the overall dataset. When the dataset size is comparatively large, such as in [21, 32], the authors have allocated data for a held-out dataset to evaluate the performance of their trained models. However when the dataset is comparatively small, like in [11, 71], the performance of the trained model is evaluated using a cross-validation technique. The evaluation metrics are discussed in detail in the next subsection. A common challenge faced by the data acquisition methods in the field is that it is time-consuming and requires expertise to assemble a large, high-quality, image dataset that can effectively train a machine learning model. Data augmentation has been widely used to overcome this challenge. Among studies reviewed in this work, some of the common augmentation techniques include cropping [20, 21, 30], rotation [18, 19, 23, 69], and translation [18]. As mentioned previously, simulated data have also been used to supplement the data from real imaging sources. Figure 4 lists the dataset characteristics observed in the studies reviewed in the current work.

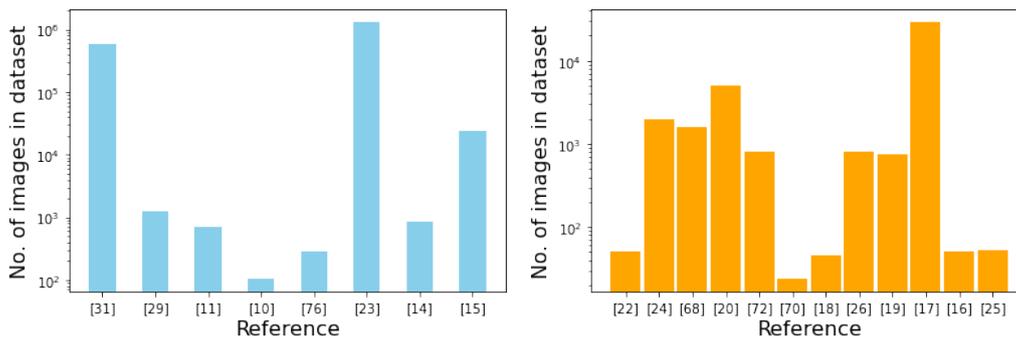


Figure 4: A representation of the dataset sizes used by a subset of studies reviewed in this work. Since coarse classification and pixel-level classification have dataset requirements spanning different orders of magnitude, they are represented in separate charts

3.3 Model selection and training

The selection of the appropriate model for a given application depends on factors such as the the problem statement, distribution and size of the dataset, the computational resources available, etc. The field of IDML for material characterization has adopted a wide variety of models, ranging from shallow to deep learning. Reflecting the current state of the field, the models can be categorized into

two distinct categories: non-neural network and neural network. The former belongs to a shallow learning paradigm, whereas the latter belongs to a deep learning paradigm. Shallow learning refers to a learning paradigm in which the model parameters (for example, the weights of support vector machine model) are learnt directly from the input features. In other words, the input features are connected directly to the output via the model. In contrast, deep learning refers to a learning paradigm in which the input features are sequentially fed to one or more intermediate layers (also known as hidden layers) before an output is obtained. In this paradigm, only the first intermediate layer learns directly from the input features. The hidden layers learn an alternate, often lower-dimensional, representation of the dataset. Owing to model simplicity and the fewer number of trainable parameters, shallow learning models are easy to train and can provide reasonable performance with a small training dataset. However, the training process is strictly empirical and it is less likely to learn hidden patterns in the dataset. On the contrary, a deep learning method is computationally expensive to train and requires a relatively larger dataset for training compared to shallow learning models. However, the trained hidden layers can provide an insight into the underlying patterns in the dataset, and can be used to explain the predictions of the model.

Coarse-grained classification tasks that generate one class label per image have been demonstrated with shallow learning algorithms such as Support Vector Machine [11,15,16,77] and Random Forest Classifier [11,31]. Reflecting the growing popularity of the convolutional neural network (CNN) in related fields, there has been a push in the application of CNNs for analyzing microstructure images. This trend can be attributed to two main factors: improved performance for multi-label classification and its application for semantic segmentation. It is important to note that CNNs and deep CNNs (DCNNs) are often selected, as opposed to multi-layered perceptron networks (MLPs) because the image data input is conducive to CNNs, which requires a spatially-dependent data format [94]. CNNs and DCNNs take input image data and convolve intermediate image data with learned kernels in several successive layers, allowing for the network to learn highly nonlinear features.

The design of a typical CNN is such that it learns about inherent patterns, through trainable filters, at different spatial resolutions compared to the input image. Libraries, such as Keras [65], enable a visualization of these patterns [30] and thus can provide insight into the rationale behind the model’s predictions. A few examples that perform coarse classification using CNNs include studies detailed in [24,30,32]. The rising popularity of deep learning techniques for semantic segmentation applications in allied fields [97] is reflected in an increased adoption of neural networks for pixel-based classification of material images. It should be noted a typical neural network architecture for coarse classification is different from a typical neural network architecture for pixel-level classification. For the former, the neural network (CNN) progressively downsamples the input image and provides as output a vector of class probabilities. The part of the model that is responsible for downsampling an input image is typically known as an encoder. For a pixel-level classification, the encoder is connected to a decoder which upsamples the lower resolution image to provide an output of the original resolution. Examples of studies that have used an encoder-decoder structure include [21,23,70–73]. The emerging field of semantic segmentation is discussed in a later section. Notable advancements in neural network architecture that have been adopted by the material science community include the use of Mask R-CNN for instance segmentation [70], use of an architecture with atrous convolution layers [74,75], use of skip connections [26], etc. A few examples of model architectures used in the field have been highlighted in Figure 5. The availability of trainable filters have also helped in augmenting domain-specific knowledge into the training process [95,96].

Depending on the availability of data, two different training modes can be implemented. In the first mode, the model is trained from a random initial state, i.e., the parameters of the model such as the filter weights are initialized to a specified random distribution. During the training period, these parameters converge to values that are optimal for the dataset. The number of training samples for this mode scales with the number of parameters, and hence it may not be feasible to implement complex deep learning models from scratch to data-deficient applications. Previously implemented works which have performed training from scratch include [25,30,69,85]. The second training mode is referred to as transfer learning, and takes advantage of models that have been pre-trained on large datasets. In this mode, the model parameters are initialized to the values that were obtained at the end of the pre-training. Subsequently, a part or whole of the model is fine-tuned with the dataset in hand. It is not common to directly use the pre-trained model parameters without fine-tuning, unless the current dataset and the dataset for pre-training are similar in nature. A major advantage of the transfer learning is that complex deep learning models can be implemented using comparatively fewer training data. Previously implemented works which have performed transfer learning include [20,23,71,73]. An emerging sub-field within transfer learning is few-shot

learning [129], in which a pre-trained model is fine-tuned using a very small training dataset (often 5 images). Few shot learning in particular, and transfer learning in general, is especially beneficial in cases where images dissimilar to the training dataset are encountered. In addition, transfer learning also reduces the demand for computational resources imposed by large models, and hence is likely to help in implementing machine learning at the edge (for example, in individual computers connected to microscopes). The reader can refer to open source libraries such as [128] which have made it easy to load pre-trained states of commonly used models.

In some cases, the bottleneck may be in the generation of ground truth data for supervised training. One way to overcome this is to explore the possibility of unsupervised training, i.e., training without the presence of ground truth labels. Unsupervised training enables models to learn to separate the data into distinct classes without explicitly learning what these classes represent. This method is particularly useful for cases where it is difficult to ascertain the number of classes beforehand. Examples of models based on unsupervised learning can be found in [29, 130, 131].

Table 2 lists the classification models and image processing algorithms employed by the 20 papers reviewed in this study.

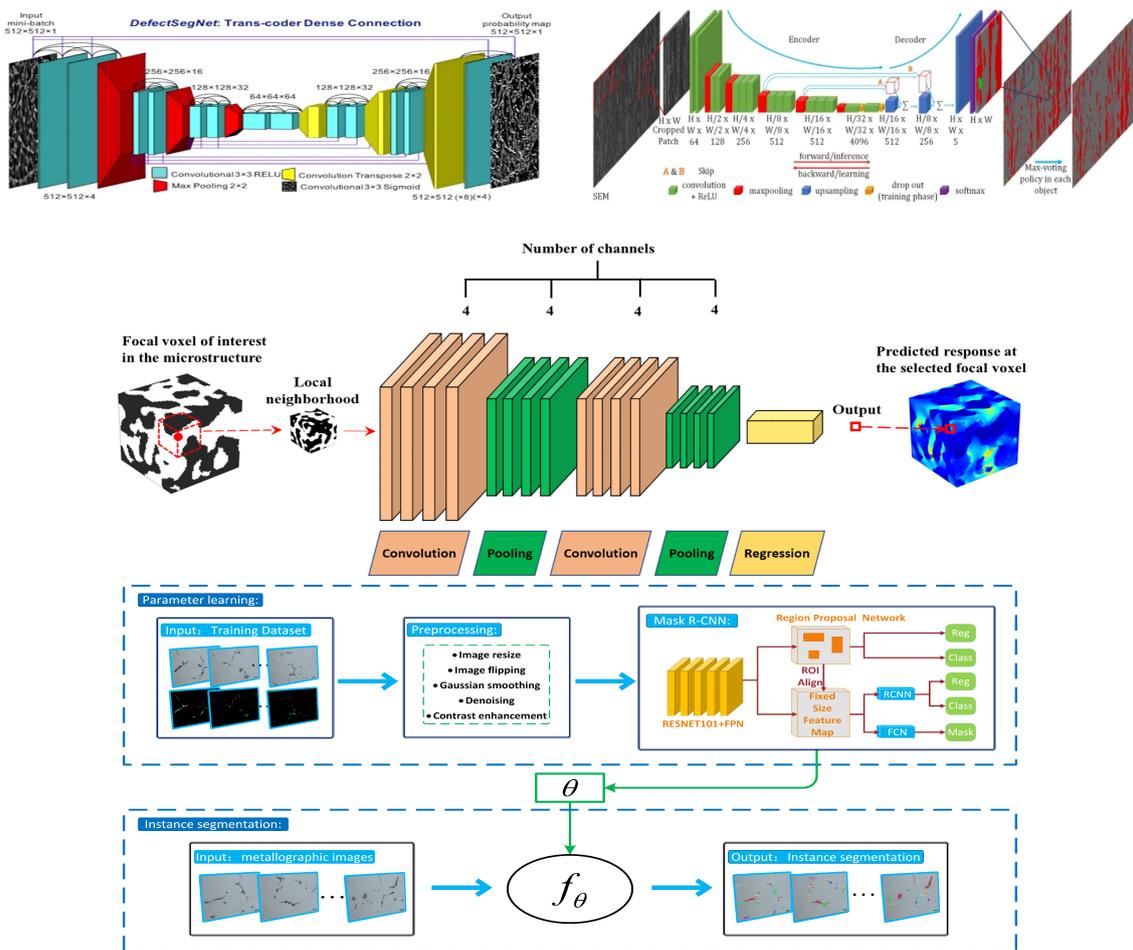


Figure 5: A selection of models employed by studies reviewed in this work. From top left, the figure were reproduced with permission from [26]; Creative Commons CC BY, [21]; Creative Commons CC BY, [85]; Copyright Elsevier (2019), and [70]; Creative Common CC BY.

3.4 Model evaluation

Choosing the appropriate evaluation metric is important in order to obtain a fair assessment of the trained model. A variety of evaluation metrics can be seen among the published literature in the field. The choice of a metric is dependent on factors such as the domain-specific problem, size of the dataset, distribution of data among the different candidate classes, etc. For example, pixel-based classification studies incorporate a distinct set of metrics to evaluate the segmentation performance of different classes. It is unfair to compare the performance of two different IDML models with similar domain-specific problem statements, if the models are implemented on different

Table 2: The class of model and their corresponding training process implemented in a subset of studies reviewed in this work.

Model used \ Training type	SVM	Neural network	Others
Training from scratch	[10, 11, 15, 16, 77]	[14, 17–19, 24–27, 30, 32, 69, 70, 72, 75, 85, 86, 88, 90–92]	[11, 29, 89]
Transfer learning	NIL	[11, 20, 21, 23, 71, 73, 74]	NIL

datasets or if they are evaluated using different metrics. However, it is useful to understand the different metrics employed by the studies and the rationale behind the choice. Some of the common evaluation metrics are listed below and are defined based on the following statistics:

- True positives, TP_i , are data of class label i that were correctly classified as i ;
- True negatives, TN_i , are data not belonging to class label i that were not classified as i ;
- False positives, FP_i , are data not belonging to class label i that were incorrectly classified as i ; and
- False negatives, FN_i , are data of class label i that were not classified as i .
- For binary classification, average accuracy is defined as

$$\frac{TP + TN}{TP + TN + FP + FN}. \quad (1)$$

This is the most common evaluation metric used, and provides an overall effectiveness of the classifier. This metric has been used for coarse classification ([11, 15, 24, 30, 32, 77]) as well as for pixel-level classification ([17, 18, 26, 70]). The average accuracy is an appropriate metric when data is uniformly distributed across class labels.

- For binary classification, precision (P) and recall (R) are defined as follows:

$$P = \frac{TP}{(TP + FP)} \quad (2)$$

and

$$R = \frac{TP}{(TP + FN)} \quad (3)$$

For multi-class classification, two different ways of averaging are followed: micro-averaging and macro-averaging. Micro-averaged precision and recall are defined as the application of the respective metric on the cumulative sum of TP, TN, FP, and FN over all the individual classes:

$$P_\mu = \frac{\sum_i TP_i}{\sum_i (TP_i + FP_i)} \quad (4)$$

Macro-averaged precision and recall are defined as the average of the respective metric applied to each of the individual classes:

$$R_M = \frac{\sum_i R_i}{l} \quad (5)$$

where l is the number of classes. Precision and recall are more appropriate metrics than average accuracy when there is an imbalance of data points across different classes. For example, in a pixel-level classification to segment defects, the number of defect pixels is likely to be small compared to the number of non-defect pixels or background pixels. Examples of studies that have used these two metrics include [15, 26, 69–71, 73].

- The F-score metric, F_{score} , derives from the precision and recall values and is defined as follows:

$$\frac{(\beta^2 + 1)PR}{(\beta^2 P + R)} \quad (6)$$

A common value that is used for β is 1, and the corresponding F1 score is the geometric mean of the precision and recall. Examples of studies that have used the F1 metric include [17, 20, 69, 70]. The F score is important towards in designing a model with a specific trade-off between precision and recall.

- Intersection over Union (IoU): This metric is used in applications of object or phase segmentation. When segmentation is performed through pixel-level classification, IoU can be defined as follows:

$$IoU = \frac{TP}{(TP + FP + FN)}. \quad (7)$$

Examples of studies that have used this metric include [17, 20, 21, 26, 27, 70, 71].

- Other metrics: Some of the studies reviewed in this work have employed evaluation metrics that are unique to the specific problem at hand. While the use of such metrics make it difficult to compare the model’s performance with similar work, they are a useful measure of the effectiveness of IDML in solving the material science problem. Examples of such metrics include difference in predicted and actual interphase connectivity [18], relative errors in the predicted center of mass and volume of grains as compared to the ground truth values [72], comparison of synthesized and original 3-dimensional structures based on spatial correlation metrics such as two-point correlation function and cluster function [14], etc.

Table 3 lists the evaluation criteria used by each of the 20 papers reviewed in this study.

Table 3: Evaluation methods adopted by a subset of studies reviewed in this work.

Evaluation metric	References
Average accuracy	[10, 11, 15–18, 24–26, 30, 32, 70, 75, 77, 91]
Precision and Recall	[15, 26, 69–71, 73, 92]
F1	[17, 20, 69, 70]
IoU	[17, 20, 21, 26, 27, 70, 71, 74]
Material specific metrics	[14, 22, 23, 29, 72, 74, 85, 86, 88–90]

3.5 Integration with the existing workflow

The final step in leveraging the capability of artificial intelligence to improve materials characterization is integrating the analysis with existing instrumentation or simulation workflows, and thereby handling the implementation challenges. Most of the studies reviewed in the work have focused on implementing state of the art machine learning or computer vision techniques to a representative dataset of images that exemplify their domain of study. In contrast, examples of integration of IDML with existing workflows can be seen in the works of Gobert et al. [15] and Scime et al. [16]. Both these studies have incorporated an imaging source to their additive manufacturing infrastructure in order to perform in-situ analysis of melt pools. In general, the integration of such techniques with existing workflows present a different set of challenges as compared to the process of adoption of IDML to material datasets. For example, the domain of electron microscopy is fast emerging as multi-modal and data-intensive [98]. The use of high-performance computing infrastructure could enable the speed of analysis to keep pace with the data generation, and also augment the efforts to automate data acquisition through real-time control of instrumentation parameters. As an example, Patton et al. [99] demonstrate the implementation of a deep learning network on a GPU supercomputer to extract structural information from atomic-resolution microscopy data. A software pipeline, built over a supercomputing architecture, to perform segmentation of three-dimensional electron microscopy data directly from the source is discussed in Vescovi et al. [100]. Seal et al. [101] report the work in progress towards the use of high-performance computing for segmenting large-scale images. Factors such as an increasing access to powerful computing resources as well as a focus on developing models that can be run at the edge, i.e., connected to existing instrumentation capabilities, are likely to accelerate the integration.

4 Challenges with IDML in the Materials Science Domain

Although there are several advantages of using an IDML approach in many materials science studies, there remain several challenges that are crucial to the growth and more widespread application of IDML. Among them, small or imbalanced data sets, noisy data, or standard validation techniques are prominent. In addition, there are very limited benchmark data sets widely available for the purpose of comparing different models.

In part, the rapid growth in the machine learning in materials science field has contributed to these challenges, and made it difficult to understand/identify appropriate methods for domain-specific goals, best practices, and opportunities [102]. To address some of these areas, emerging techniques, best practices, and potential new paths are subsequently discussed.

5 Emerging techniques and potential new paths

Microstructure characterization using methods from ML and AI has been a nascent endeavour as indicated in the previous sections. The revolution of deep learning to solve seemingly complex computer vision tasks like image recognition has fueled researchers in material science to transfer some of these successes to the field. Section 3 has represented a broad overview of this work from the perspective of application of state of the art machine learning technologies to materials characterization. A lot of this work involves using image datasets of microstructures and other materials characteristics to solve controlled classification problems. Even though this represents a major advancement in the field, several problems still remain as discussed in Section 4. These challenges can be addressed by emerging methodologies in AI that have the potential to alleviate some of the concerns raised in this study.

One of the challenges in materials characterization is the apparent dearth of labelled datasets. Deep learning problems require massive amounts of labelled data to produce results. Unfortunately, annotating data is time consuming and expensive. Recent advancements in AI promise methods to tackle this problem such as active learning. Active learning is inspired from the idea that a predictor that is trained on a smaller set of well-chosen samples can perform as efficiently as a predictor that is trained on a much larger number of randomly sampled data points. In particular, active learning can be implemented as a form of expert-machine interaction. Starting from a small non-optimally sampled batch of data, the algorithm presents the users the images or pixels whose inclusion in the set improves the performance of the predictor. The user or domain expert interacts with the machine in the process of annotation. The procedure is iterated until a stopping criteria has been achieved. Approaches for active learning have been demonstrated in image classification [103], deep learning [104], remote sensing [105], medical imaging [106] and image retrieval [107]. A similar approach can be taken for materials characterization where active learning can be used to reduce annotation burden of material scientists.

A alternate set of methods to reducing the problem of small datasets is known as semi-supervised learning. It considers the problem of supervised learning when only a small subset of data is available for training, which is exactly the challenge that we are looking to overcome. Semi supervised learning algorithms address this problem by learning from unlabelled data as well as labelled data to build better predictors [108]. Semi supervised approaches have been explored for multi-model image classification [109] and for deep generative modelling [110]. Material characterization will improve with the help of semi-supervised learning algorithms that learn from unlabelled data which is much easier to curate.

A common problem with modern ML and DL methods is that they are inherently inscrutable. Even though they are responsible for reproducing results with astonishing accuracy, they fundamentally lack the ability to provide explanations for their predictions. As a consequence, it is very difficult to interpret the results of a classification or segmentation model. A number of approaches have been proposed recently to tackle this problem. Common techniques for ML model explainability is provided here [111]. Another way is to visualize the model [112, 113] by observing the values of filters in trained convolutional neural networks. Examples of studies which have used these techniques to explain model predictions on material characterization datasets include [135] and [134]. In [135], a CNN was used for prediction of electromagnetic response of nanophotonic structures, following which relevancy-based heatmaps were used to highlight the spatial regions in the input that were important for the prediction. Similarly, the technique of computing saliency maps was utilized in [134] to gain insight into the property prediction (short-circuit current) from morphology maps of organic photovoltaic films. Using such techniques to understand the latent representation of data and the trained state of hidden layers could accelerate the process of material design and discovery. Another viewpoint of interpretability is by reducing model complexity. The core idea is that simpler models are easier to understand than larger esoteric models. For example, [114] tries to approximate the complex model with a locally linear model. A related idea is using feature importance [115] for interpreting predictions. Symbolic approaches [116–118] generate symbolic expressions to generate semantic understanding of the predictions. An example of an interpretable workflow for material characterization is [136], in which a regression model is developed to predict peak stress from a morphology image generated from SEM. Using a generative

method, structural features (such as porosity) are systematically modified to study its effect on the stress. This analysis enables one to identify the morphological features that are key for a targeted output. As discussed before, these challenges involving interpretability need to be addressed before materials characterization using IDML is to be widely accepted.

5.1 Generative methods

Before GANs, there have been many generative models, such as the Deep Boltzmann Machine (DBM) and Variational Auto-encoder (VAE). However, they had less of an impact due to the difficulty of approximating intractable probabilistic computations in maximum likelihood estimation, known as the “maximum likelihood training paradigm”, and leveraging the benefits of piecewise linear units in the generative context [119]. Unlike previous generative models, Generative Adversarial Network (GAN) brings a new breath to the deep generative models with its simple structure and effectiveness in image generation. With the help of Convolutional Neural Networks, the GANs can generate realistic images, as large as 1024 by 1024 pixels.

A GAN framework consists of a generator, G , that generates samples from a noise variable, z , and a discriminator, D , that aims to distinguish between samples from the real data distribution and those from the synthetic data distribution (from the generator). The generator can be thought of as counterfeiters trying to produce fake but realistic images, while the discriminator aims to distinguish the fake images from the real ones.

GANs have been proven successful for many image synthesis and unsupervised learning tasks. It is a popular framework for representation learning, such as disentangle pose from lighting in 3D rendered images [120], and image completion, where large missing regions are synthesized utilizing the surrounding image features [121]. Variants of GANs have surpassed many other generative models in the quality of samples as well as their underlying representation. GANs have also been applied to the task of image-to-image translation, which learns a mapping between an input image and an output image. Several examples of this image-to-image translation exist in literature, including [122] used CycleGAN to transfer images of one style to another, such as between landscape images in summer and in winter, and between photographs and paintings of Monet. [123] used GauGAN to create photorealistic images from segmentation maps, which are labeled sketches that depict the layout of a scene. [124] used Pix2Pix on many applications of image-to-image translation, such as mapping from aerial photos to maps, and mapping from edges to photos [124].

Recently, GANs have emerged as a promising methodology for application in computational materials design, for the purpose of developing structure-property and structure-performance relations via physical simulations [31, 125, 126]. With the help of CNN backbone, GANs are able to capture complex microstructural characteristics. In [133], GAN is used to generate high-resolution three-dimensional images of the pore space at different scales. In [125], the authors proposed a deep adversarial learning methodology to overcome the limitations of existing microstructure characterization and reconstruction (MCR) techniques. In this work, the GAN is used identify the mapping between latent codes and microstructures. In [126], the GANs are used to learn a mapping between latent variables and microstructures, which is later used as the design variables to obtain microstructures of desired material property with the help of a Bayesian optimization framework. In [31], the authors generated high-resolution microstructure images with the advanced progressive-growing GAN (pgGAN), and interpret the task of microstructure generation as microstructure distribution characterization and image-to-image translation. Figure 6 highlights the two types of applications that have benefited from using generative models.

5.2 Merging data from multiple characterization techniques

Materials characterization studies typically involve the application of multiple experimental techniques, sometimes in combination with modelling or computation, in order to develop microstructure-processing-property relationships. The multi-length scale nature of such studies necessitates varying data types to characterize any microstructure, which is schematically shown in Figure 7. The multiple data types shown here span from atom probe tomography data (sub-nanometer scale spatial resolution, typically analyzing volumes to 50 nm by 50 nm by 100 nm) to TEM images, to SEM and optical micrographs. These data types are only a few examples of the data types used in materials characterization. Considering a more specific example, of analyzing an alloy after casting. This alloy may then be subjected to a homogenization anneal, and subsequently several thermo-mechanical processing to achieve the desired dimensions and properties. To examine the microstructures produced after processing, a researcher may begin with cross-sectioning

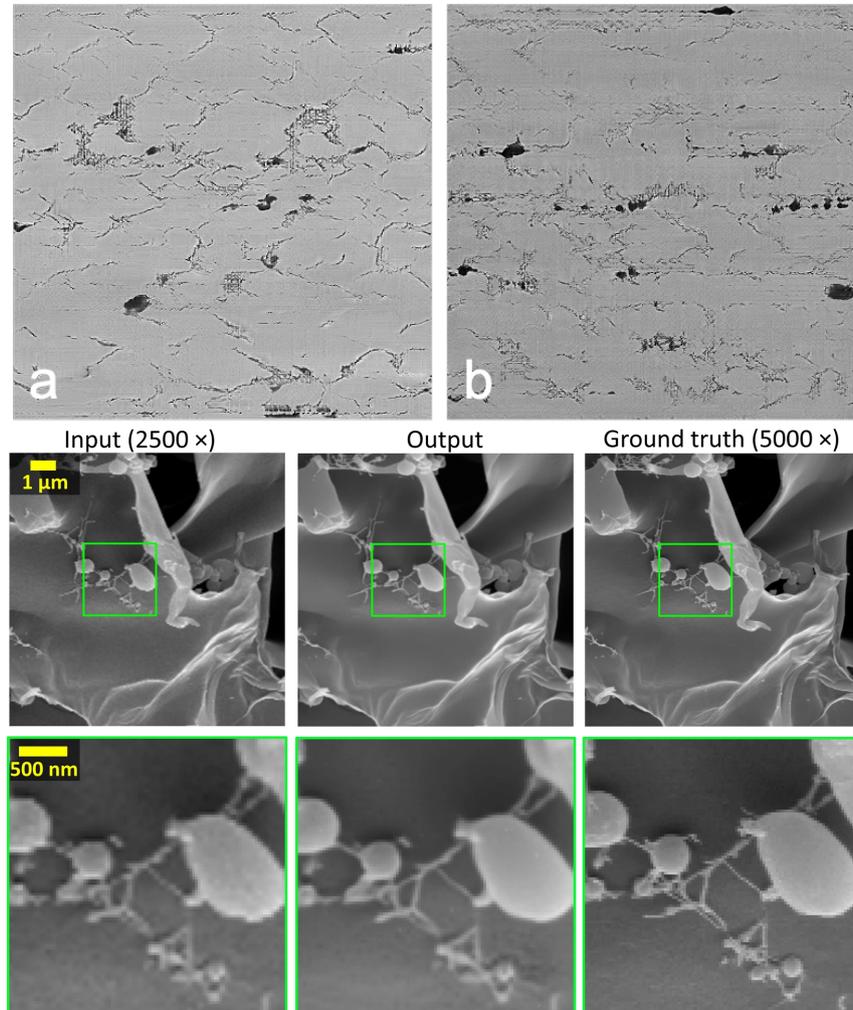


Figure 6: Two types of applications in the field of materials characterization that have benefited from using generative models. (Top) Example synthetic images generated by the trained Progressive Growing GAN. The generated images show varying microstructural features, specifically different extent of lamellar transformation products, and distribution of carbides. Figure reproduced with permission from [31]; Creative Common CC BY. (Bottom) The use of a generative model to enhance the resolution of an image acquired from a Scanning Transmission Electron Microscope. The benefit of this application lies in a reduced imaging time as well as a reduced likelihood for electron beam damage to the samples. Figure reproduced with permission from [90]; Creative Commons Attribution 4.0 International.

and polishing the material and examining the cross-section using an optical microscope. Some microstructural features may be visible, depending on the microstructure (i.e. grain boundaries, inclusions, etc.), but others may not, hence the researcher will go to the SEM to look at the same microstructure with higher resolution. Secondary electron (SE) and back scatter electron (BSE) images may be collected, all images, but providing different detail about the microstructure (i.e. topography for SE images versus atomic number contrast for BSE images). Energy dispersive spectroscopy (EDS) may be performed to obtain element distributions or composition information from varying regions of the microstructure. EDS data can be collected in points, or as maps. Now, multiple image and data types have been collected, all providing unique detail about the microstructure in question. Next, electron backscatter diffraction (EBSD) may be collected to provide grain orientation information, and another data type and instance is provided. For even higher resolution characterization, transmission electron microscopy (TEM) may be employed to identify structure, and perform EDS mapping and point analysis in small volumes, with a typical sample size of $10\mu\text{m}$ by $10\mu\text{m}$ with less than 100 nm thickness (to be electron transparent). Higher resolution chemical and isotopic mapping can also be performed using atom probe tomography (APT), yielding a three dimensional point cloud data set with the x,y,z position and elemental and isotopic identity of each atom detected. The workflow detailed here is just one example, with limited techniques. Several other techniques exist and are selected based on the material of interest and characterization challenge faced. The methods discussed here are limited to characterization of a material via microscopy methods that yield data primarily in the form of images, without regard to thermal or mechanical properties. Yet, the key concept demonstrated through this example is sound, and remains a critical component to the future of machine learning in materials characterization and design: a single material condition can be described by multi-lengthscale, multi-modal data sets. While many IDML research works have been performed using images from a single imaging modality, the concept of fusing multiple imaging modalities is still in its infancy in materials science. Multimodal learning is when learning algorithms relate information from multiple data collection sources, as detailed by Ngiam et al [138], who focused on relating speech audio to videos of lip movement. In the materials science domain, a multimodal learning problem could involve data collected using multiple different chemical imaging modalities in order to relate information such as crystal structure to composition, for example. Independent of application area, multimodal data fusion aims to capture correlations across modalities, such as from SEM images that detail microstructural features of interest to TEM and APT that detail structure and composition of phases present, grain boundary character, etc. Multimodal chemical imaging, correlative and complementary microscopy experiments are uniquely powerful since they allow for correlation of properties with chemical composition and structure, which is central to understanding how these drive material performance [137].

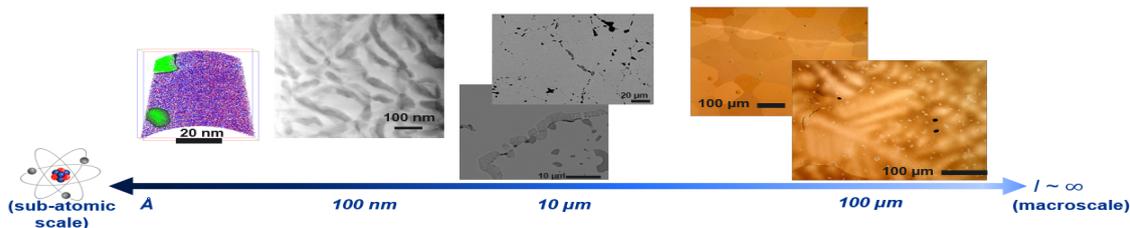


Figure 7: Schematic of data from different image types that span across length scales and are instrumental to materials characterization studies. From left to right: atom probe tomography data, STEM bright field image, SEM micrographs, optical micrographs.

6 Summary

In this work, we have reviewed the field of image driven machine learning (IDML) towards the analysis of material structure characterization. The scope of the discussion included studies based on sample surface images from cameras, digital images from optical and electron microscopes, diffraction pattern images, and simulated images of material structures at various length scales. A canonical series of six action steps was defined, recognizing the importance of multiple independent stages required for a successful application of IDML in the field: problem definition, dataset building, model selection, model training, model evaluation, and integration with the existing workflow. Two of the most widely adopted characterization problems have been morphology class detection

using image classification (one class label per image) and phase segmentation for quantitative analysis using pixel-level classification (one label per pixel). An emerging sub-field in the last couple of years has been the adoption of generative models to augment the image acquisition process, especially useful for high-resolution and destructive characterization techniques such as electron microscopy. In the last few years, the field has seen the widespread adoption of neural network models such as the convolutional neural network for image classification as well as for semantic segmentation through pixel-level classification. Though the difficulty in acquiring high quality data in the field is a challenge to training complex models, the rise in adoption of the training paradigm of transfer learning promises to partially alleviate this limitation. In large part, the material science community has thus far shown the utility of image driven machine learning as a useful paradigm to characterize material structures. With increasing access to high-performance computing and data architectures, robust model development through the incorporation of interpretability and explainability, the next stage of progress for the field will entail the integration of IDML into existing instrumentation and simulation workflows.

7 Acknowledgements

This work was in part supported by the U.S. Department of Energy (DOE) National Nuclear Security Administration, and in part supported by the National Science Foundation ; Project award No. CMMI-1729336. A portion of this work was performed at the Pacific Northwest National Laboratory (PNNL), which is operated for the U.S. DOE by Battelle Memorial Institute under Contract No. DE-AC05-76RLO1830.

References

- [1] J.Alkemper, P.Voorhees, *J. Microsc.*, **201** (2001). <https://doi.org/10.1046/j.1365-2818.2001.00832.x>
- [2] A.Devaraj, D.E. Perea, J.Liu, L. M. Gordon, T. J. Prosa, P. Parikh, D. R. Diercks, S. Meher, R. P. Kolli, Y. S. Meng, and S. Thevuthasan., *Int. Mater. Rev.* **63** (2018). <https://doi.org/10.1080/09506608.2016.1270728>
- [3] L.Luo, L.Li, D.K. Schreiber, Y. He, D. R. Baer, S. M. Bruemmer, and C. Wang., *Sci. Adv.*, **6**(17) (2020). <http://doi.org/10.1126/sciadv.aay8491>
- [4] E.J. Kautz, S.V. Lambeets, D.E. Perea, A.Y.Gerard, J.Han, J.R.Scully, J.E.Saal, and D.K.Schreiber, *Scr. Mater.*, **194** (2021). <https://doi.org/10.1016/j.scriptamat.2020.10.051>
- [5] D.L. McDowell, S.R. Kalidindi, *MRS Bull.*, **41**(4) (2016). <https://doi.org/10.1557/mrs.2016.61>
- [6] A. Agrawal, A. Choudhary, *APL Mater.*, **4**(5) (2016). <http://doi.org/10.1063/1.4946894>
- [7] K. Schwab, *The Fourth Industrial Revolution — Foreign Affairs*, <http://www.foreignaffairs.com/articles/2015-12-12/fourth-industrial-revolution>. Accessed on Dec 20, 2020
- [8] J.M. Rickman, T.Lookman, S.V. Kalinin, *Acta Mater.*, **168**(473) (2019). <http://doi.org/10.1016/j.actamat.2019.01.051>
- [9] R.Ramprasad, R.Batra, G.Pilania, A. Mannodi-Kanakithodi, and C.Kim, *npj Comput. Mater.*, **54** (2017). <http://doi.org/10.1038/s41524-017-0056-5>
- [10] B.L. DeCost, E.A. Holm, *Comput. Mater. Sci.*, **110**, (2015). <http://doi.org/10.1016/j.commatsci.2015.08.011>
- [11] A.Chowdhury, E.Kautz, B.Yener, D.Lewis, *Comput. Mater. Sci.*, **123**, 176 (2016). <http://doi.org/10.1016/j.commatsci.2016.05.034>
- [12] C. S. Smith, “A History of Metallography”, University of Chicago Press, 1960.
- [13] H.Abrams, *Metallography*, **4**(1), 59 (1971). [https://doi.org/10.1016/0026-0800\(71\)90005-X](https://doi.org/10.1016/0026-0800(71)90005-X)
- [14] F.Zhang, Q.Tenga, H.Chen, X. He, X. Dong., *Comput. Mater. Sci.* **186** (2021). <https://doi.org/10.1016/j.commatsci.2020.110018>

- [15] C.Gobert, E.W. Reutzler, J.Petrich, A.R. Nassar, S.Phoha, *Addit. Manuf.*, **21** (517) (2018). <http://doi.org/10.1016/j.addma.2018.04.005>
- [16] L.Scime, J.Beuth, *Addit. Manuf.*, **25** (2019). <http://doi.org/10.1016/j.addma.2018.11.010>
- [17] T.Stan, Z.T. Thompson, P.W. Voorhees, *Mater. Charact.* **160** (2020).
- [18] T.Strohmann, K.Bugelnig, E.Breitbarth, F. Wilde, T. Steffens, H. Germann, G. Requena, *Sci. Rep.*, **9** (2019). <https://doi.org/10.1038/s41598-019-56008-7>
- [19] S.Evsevleev, S.Paciornik, G.Bruno, *Adv. Eng. Mater.*, **22**(4) (2020). <https://doi.org/10.1002/adem.201901197>
- [20] S.Tsopanidis, R.H. Morenz, S.Osovski, *Eng. Fract. Mech.* **231** (2020)
- [21] S.M. Azimi, D.Britz, M.Engstler, M. Fritz, F. Mucklich, *Sci. Rep.*, **8** (2018). <http://doi.org/10.1038/s41598-018-20037-5>
- [22] A.Campbell, P.Murray, E.Yakushina, S.Marshall, W. Ion, *Mater. Des.*, **141** 395 (2018). <https://doi.org/10.1016/j.matdes.2017.12.049>
- [23] R.Agbozo, W.Jin, *J. Korean Soc. Precis. Eng.*, **37**(5) 361 (2019).
- [24] G.D. Forster, A.Castan, A.Loiseau, J. Nelayah, D. Alloyeau, F. Fossard, C. Bichara, H. Amara, *Carbon*, **169**, 465 (2020). <https://doi.org/10.1016/j.carbon.2020.06.086>
- [25] M.Ziatdinov, O.Dyck, A.Maksov, X. Li, X. Sang, K. Xiao, R. R. Unocic, R. Vasudevan, S. Jesse and S. V. Kalinin, *ACS Nano*, **11** (2017) <https://doi.org/10.1021/acsnano.7b07504>
- [26] G.Roberts, S.Y. Haile, R.Sainju, D. J. Edwards, B. Hutchinson, Y. Zhu, *Sci. Rep.*, **9**(1) (2019). <https://doi.org/10.1038/s41598-019-49105-0>
- [27] L.Yao, Z.Ou, B.Luo, C. Xu, and Q. Chen, *ACS Cent. Sci.*, **6** (2020). <https://doi.org/10.1021/acscentsci.0c00430>
- [28] Y.Weil, Z.Peng, M.Kuhbach, A. Breen, M. Legros, M. Larranaga, F. Mompiau, B. Gault, *PLoS One*, **14**(11) (2019). <https://doi.org/10.1371/journal.pone.0225041>
- [29] H.Chan, M.Chelukara, T.D. Loeffler, B. Narayanan, and S. K. R. S. Sankaranarayanan, *npj Comput. Mater.*, **6**(1) (2020). <https://doi.org/10.1038/s41524-019-0267-z>
- [30] A.Baskaran, G.Kane, K.Biggs, R. Hull, D.Lewis, *Comput. Mater. Sci.*, **177** (2020). <https://doi.org/10.1016/j.commatsci.2020.109593>
- [31] W.Ma, E.J. Kautz, A.Baskaran, A. Chowdhury, V. Joshi, B. Yener, and D. J. Lewis, *J. Appl. Phys.*, **128**(13) (2020). <https://doi.org/10.1063/5.0013720>
- [32] J.A. Aguiar, M.L. Gong, R.R. Unocic, T. Tasdizen, and B. D. Miller, *Sci. Adv.*, **5**(10) (2019). <http://doi.org/10.1126/sciadv.aaw1949>
- [33] R.Bostanabad, Y.Zhang, X.Li, T. Kearney, L. C. Brinson, D. W. Apley, W. K. Liu, W. Chen, *Prog. Mater. Sci.*, **95** (2018). <https://doi.org/10.1016/j.pmatsci.2018.01.005>
- [34] N.H. Paulson, M.W. Priddy, D.L. McDowell, S.R. Kalidindi, *Acta Mater.*, **129**, 428 (2017). <https://doi.org/10.1016/j.actamat.2017.03.009>
- [35] S.Torquato, G.Stell, *J. Chem. Phys.*, **77**(4) (1982).
- [36] S.R. Kalidindi, S.R. Niezgodna, A.A. Salem, *JOM*, **63**(4) (2011). <https://doi.org/10.1007/s11837-011-0057-7>
- [37] P.E. Chen, W.Xu, N.Chawla, Y. Ren, Y. Jiao, *Acta Mater.*, **179**, 317 (2019). <https://doi.org/10.1016/j.actamat.2019.08.045>
- [38] P. Acar and V. Sundararaghavan, *AIAA*, **55** (8), (2017)
- [39] T.Huang, J. Gao, Q. Sun, D. Zeng, X. Su, W. K. Liu, W. Chen, *Comp. Struc.*, **260**(113470) (2021)

- [40] P.Nikolaev, D.Hooper, F.Webber, R. Rao, K. Decker, M. Krein, J. Poleski, R. Barto, and B. Maruyama , *npj Comput. Mater.*, **2** (2016). <http://doi.org/10.1038/npjcompumats.2016.31>
- [41] W.Ye, C.Chen, S.Dwaraknath, A. Jain, S. P. Ong, and K. A. Persson , *MRS Bull.*, **43**(9), 664 (2018). <http://doi.org/10.1557/mrs.2018.202>
- [42] C.Oses, C.Toher, S.Curtarolo, *MRS Bull.*, **43**(9), 670 (2018). <http://doi.org/10.1557/mrs.2018.207>
- [43] C.Draxl, M.Scheffler, *MRS Bull.*, **43**(9), 676 (2018). <http://doi.org/10.1557/mrs.2018.208>
- [44] J.J. Plata, P.Nath, D.Usanmaz, J. Carrete, C. Toher, M. de Jong, M. Asta, M. Fornari, M. B. Nardelli, and S. Curtarolo, *npj Comput. Mater.*, **3**(1), 45 (2017). <https://doi.org/10.1038/s41524-017-0046-7>
- [45] S.V. Kalinin, B.G. Sumpter, R.K. Archibald, *Nat. Mater.*, **14**, (2015). <https://doi.org/10.1038/nmat4395>
- [46] E.J. Kautz, A.R. Hagen, J.M. Johns, D. E.Burkes.,*Comput. Mater. Sci.*, **161**, 107 (2019). <https://doi.org/10.1016/j.commatsci.2019.01.044>
- [47] Q.Liu, H.Wu, M.J. Paul, P. He, Z. Peng, B. Gludovatz, J. J. Kruzic, C. H. Wang, X. Li., *Acta Mater.*, **201** 316 (2020). <https://doi.org/10.1016/j.actamat.2020.10.010>
- [48] S.F.Fang, M.P.Wang, M.Song, *Mater. Des.*, **30**(7), 2460 (2009). <http://doi.org/10.1016/j.matdes.2008.10.008>
- [49] Z.Yang, Y.C.Yabansu, R Al-Bahrani, W. Liao, A. N. Choudhary, S. R. Kalidindi, A. Agrawal,*Comput. Mater. Sci.* **151**, (2018). <http://doi.org/10.1016/j.commatsci.2018.05.014>
- [50] A.Seko, K.Toyoura, S.Muto, T. Mizoguchi, and S. Broderick , *MRS Bull.*, **43**(9), 690 (2018). <http://doi.org/10.1557/mrs.2018.206>
- [51] A.J. Ballard, R.Das, S.Martiniani, D. Mehta, L. Sagun, J. D. Stevenson, and D. J. Wales, *Phys. Chem. Chem. Phys.*, **19**(20) (2017). <http://doi.org/10.1039/C7CP01108C>
- [52] S.Chmiela, H.E. Saucedo, K.R. Maller, A. Tkatchenko, *Nat. Commun.*, **9**(1) (2018). <http://doi.org/10.1038/s41467-018-06169-2>
- [53] K.J. Jose, N.Artrith, J.Behler, *J. Chem. Phys.*, **136**(19) (2012).
- [54] A.P. Bartok, S.De, C.Poelking, N. Bernstein, J. R Kermode, G. Csányi, M. Ceriotti, *Sci. Adv.*, **3**(12) (2017). <https://doi.org/10.1126/sciadv.1701816>
- [55] E.Kim, K.Huang, A.Saunders, A. McCallum, G. Ceder, E. Olivetti, *Chem. Mater.* , **29**(21) (2017). <https://doi.org/10.1021/acs.chemmater.7b03500>
- [56] J.Ling, M.Hutchinson, E.Antonio, B. DeCost, E. A.Holm, B. Meredig, *Mat. Disc.*, **10** (2017). <http://doi.org/10.1016/j.md.2018.03.002>
- [57] E.Kautz, W.Ma, S.Jana, A. Devaraj, V. Joshi, B. Yener, D. Lewis, *Mater. Charact.*, **166** (2020). <https://doi.org/10.1016/j.matchar.2020.110379>
- [58] W.B. Park, J.Chung, J.Jung, K. Sohn, S. P. Singh, M. Pyo, N. Shin, K-S Sohn, *IUCrJ*, **4**(4) (2017). <https://doi.org/10.1107/S205225251700714X>
- [59] Y.Liu, T.Zhao, W.Ju, S. Shi, *J Materiomics.*, **3**(3) (2017). <http://doi.org/10.1016/j.jmat.2017.08.002>
- [60] K.T. Butler, D.W. Davies, H.Cartwright, O. Isayev, A.Walsh, *Nature*, **559**(7715) (2018). <http://doi.org/10.1038/s41586-018-0337-2>
- [61] B.M. Lake, R.Salakhutdinov, J.B. Tenenbaum, *Science*, **350**(6266) (2015). <http://doi.org/10.1126/science.aab3050>
- [62] A.Krizhevsky, I.Sutskever, G.E. Hinton, *Adv. in Neu. Inf. Proc. Sys*, **25**
- [63] J.Deng, W.Dong, R.Socher, L-J. Li, K. Li, L. Fei-Fei , in *In CVPR* (2009)

- [64] E.K.Mace, J.D.Ward, C.E.Aalseth, *J. Radioanal. Nucl. Chem.*, **318** (2018). <http://doi.org/10.1007/s10967-018-5983-1>
- [65] keras-team/keras: Deep Learning for humans. <https://github.com/fchollet/keras>. Accessed on Dec 20, 2020
- [66] A. Paszke, S. Gross, F. Massa, A. Lerer, J.B. et al., arXiv:1912.01703
- [67] M.Abadi, A.Agarwal, P.Barham, et al., <http://arXiv:1603.04467>
- [68] K.Tsutsui, H.Terasaki, T.Maemura, K. Hayashi, K. Moriguchi, S. Morito, *Comput. Mater. Sci.* **159** (2019). <https://doi.org/10.1016/j.commatsci.2018.12.003>.
- [69] Y.Zhu, Q.Ouyang, Y.Mao, *BMC Bioinf.*, **18**(348) (2017).
- [70] D.Chen, D.Guo, S.Liu, F.Liu, *Symmetry*, **12**, 639 (2020)
- [71] T.F. B.L.DeCost, E.Holm, *Microsc. Microanal.*, **25** 21 (2019)
- [72] O.Furat, M.Wang, M.Neumann, L. Petrich, M. Weber, C. E. Krill and V. Schmidt, *Front. Mater.*, **6** (2019). <https://doi.org/10.3389/fmats.2019.00145>
- [73] A.O. Vuola, S.U. Akram, J.Kannala, in *IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)* (2019), pp. 208-212
- [74] H.Hwang, S.M. Choi, J.Oh, S-M Bae, J-H. Lee, J-P. Ahn, J-O. Lee, K-S. An, Y. Yoon, J-H. Hwang, *J. Power Sources*, **471** (2020). <https://doi.org/10.1016/j.jpowsour.2020.228458>
- [75] B.Ma, X.Ban, H.Huang, Y. Chen, W. Liu, Y. Zhi, *Symmetry*, **10**(4) (2018). <https://doi.org/10.3390/sym10040107>
- [76] H.Wang, *IOP Conf. Ser.: Mater. Sci. Eng.* **652** (2019)
- [77] B.L. DeCost, H.Jain, A.D. Rollett, E.A.Holm, *JOM* **69**(3), 456 (2017). <https://doi.org/10.1007/s11837-016-2226-1>
- [78] F.Burger, C.Buck, J.Pauli, W.Luther, *Intl. Conf. on Comp. Vis. Theory and App.*, (VISAPP) pp. 143–152 (2014)
- [79] R.Lorenzoni, I.Curosu, S.Paciornik, V. Mechtcherine, M. Oppermann, F. Silva, *Cem. Concr. Compos.*, **108** (2020). <https://doi.org/10.1016/j.cemconcomp.2020.103551>
- [80] L.YiHao, H.ZiHeng, S.ZhiGuang, et al., *Sci. China: Technol. Sci.*, **62**(4) 521 (2019)
- [81] N.Lubbers, T.Lookman, K.Barros, *Phys. Rev. E*, **96** (2017). <https://doi.org/10.1103/PhysRevE.96.052111>
- [82] S.R. Niezgodna, A.K. Kanjarla, S.R. Kalidindi, *Integr. Mater. Manuf. Innov.*, **2**(1) (2013). <https://doi.org/10.1186/2193-9772-2-3>
- [83] H.Xu, D.A. Dikin, C.Burkhart, W.Chen, *Comput. Mater. Sci.*, **85** (2014). <https://doi.org/10.1016/j.commatsci.2013.12.046>
- [84] H.Xu, R.Lu, A.Choudhary, W.Chen, *J. Mech. Des.*, **137** (2015). <https://doi.org/10.1115/1.4029768>
- [85] Z.Yang, Y.C. Yabansu, D.Jha, W. Liao, A. N. Choudhary, S. R. Kalidindi, A. Agrawal,, *Acta Mater.*, **166** (2019). <https://doi.org/10.1016/j.actamat.2018.12.045>
- [86] R.Cang, M.Y. Ren, *ASME 2016 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference*, (2016). <http://doi.org/10.1115/DETC2016-59404>
- [87] I.Arganda-Carreras, V.Kaynig, C.Rueden, K. W. Eliceiri, J. Schindelin, A. Cardona, H.S. Seung, *Bioinformatics*, **33** (2017). <http://doi.org/10.1093/bioinformatics/btx180>
- [88] D.Cirean, A.Giusti, L.M. Gambardella, J.Schmidhuber, *Proc. Neural Inf. Proc. Sys.* **25** (2012)
- [89] S.Wang, G.Cao, B.Weil, Y.Yin, G.Yang, C.Li *Biomed. Eng. Online*, **12** (2013). <https://doi.org/10.1186/1475-925X-12-59>

- [90] K.de Haan, Z.S. Ballard, Y.Rivenson, Y.Wu, A.Ozcan, *Sci. Rep.*, **9**(12050) (2019). <https://doi.org/10.1038/s41598-019-48444-2>
- [91] K.Kaufmann, C.Zhu, A.S. Rosengarten, D. Maryanovsky, T. J. Harrington, E. Marin, K. S. Vecchio, *Science*, **367** (2020). <http://doi.org/10.1126/science.aay3062>
- [92] J.Madsen, P.Liu, J.Kling, J. B. Wagner, T. W. Hansen, O. Winther, J. Schiøtz, *Adv. Theor. Simul.*, **1**(8) (2018). <https://doi.org/10.1002/adts.201800037>
- [93] Y.Han, R.J. Griffiths, H.Z. Yu, Y.Zhu, *J. Mater. Res.*, **35**(15) (2020). <http://doi.org/10.1557/jmr.2020.120>
- [94] I.Goodfellow, Y.Bengio, A.Courville, *Deep Learning* (MIT Press, 2016)
- [95] B.Chidester, T.Zhou, M.N. Do, J.Ma, *Bioinformatics*, **35** (2019). <https://doi.org/10.1093/bioinformatics/btz353>
- [96] D.Marcos, M.Volpi, D.Tuia, in *23rd International Conference on Pattern Recognition (ICPR)* (2016)
- [97] A.Garcia-Garcia, S.Orts-Escolano, S.Oprea, V. Villena-Martinez, J. Garcia-Rodriguez, arXiv:1704.06857 (2017)
- [98] S. Spurgeon, C. Ophus, L. Jonestextit, et al., *Nat. Mater.*, **20** (2020) <https://doi.org/10.1038/s41563-020-00833-z>
- [99] R. M. Patton; J. T. Johnston, S. R. Young, C.D. Schuman, D.D. March, T.E, Potok, D.C. Rose, S-H.Lim, T.P. Karnowski, M.A. Ziatdinov, S.V. Kalinin, *SC18: Intl. Conf. for High Perf. Comp., Networking, Storage and Analysis*, (IEEE, 2018) <http://doi.org/10.1109/SC.2018.00053>
- [100] R.Vescovi, H.Li, J.Kinnison, M. Keceli, M. Salim, N. Kasthuri, T. D. Uram, N. Ferrier, arXiv:2011.03204 (2020)
- [101] S.K.Seal, S. Lim, D.Wang, J.Hinkle, D.Lunga, A.Tsaris, in *49th Intl. Conf. on Parallel Proc. - ICPP*, (ACM, 2020). <https://doi.org/10.1145/3404397.3404468>
- [102] D.Morgan, R.Jacobs, *Annu. Rev. Mater. Res.*, **50**(1) 71 (2020). <http://doi.org/10.1146/annurev-matsci-070218-010015>
- [103] A.J. Joshi, F.Porikli, N.Papanikolopoulos, in *2009 IEEE Conf. on Comp. Vis. and Pat. Recog.* (2009)
- [104] K.Wang, D.Zhang, Y.Li, et al., *IEEE Transactions on Circuits and Systems for Video Technology*, **27**(12) 2591 (2016)
- [105] D.Tuia, F.Ratle, F.Pacifici, M. Kanevski, and W. J. Emery, *IEEE Trans. Geosci. Electron.*, **47**(7) 2218 (2009)
- [106] A.Chowdhury, S.K. Biswas, S.Bianco, (2017). <https://doi.org/10.1101/211060>
- [107] S.Tong, E.Chang, in *Proceed. of the ninth ACM intl. conf. on multimedia* (2001)
- [108] X.J. Zhu, *Tech. rep.*. <http://digital.library.wisc.edu/1793/60444>. (2005)
- [109] M.Guillaumin, J.Verbeek, C.Schmid, in *2010 IEEE Computer society conference on computer vision and pattern recognition* (IEEE, 2010), pp. 902–909
- [110] D.P. Kingma, S.Mohamed, D.Jimenez-Rezende, M. Welling, *Adv. Neural Inf. Process. Syst.* **27**, 3581 (2014)
- [111] L.H. Gilpin, D.Bau, B.Z. Yuan, A. Bajwa, M. Specter, L. Kagal, in *2018 IEEE 5th International Conference on data science and advanced analytics (DSAA)* (IEEE, 2018), pp. 80–89
- [112] Q.Zhang, Y.Nian-Wu, S.C. Zhu, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2018), pp. 8827–8836
- [113] Q.s. Zhang, S.C. Zhu, *Front. of Inf. Tech. & Elec. Eng.*, **19**(1), 27 (2018)
- [114] M.T. Ribeiro, S.Singh, C.Guestrin, in *Proc. of the 22nd ACM SIGKDD Intl. Conf. on knowledge discovery and data mining* (2016), pp. 1135–1144

- [115] S.M. Lundberg, S.I. Lee, in *Advances in neural information processing systems* (2017), pp. 4765–4774
- [116] A.Lazaridou, A.Peysakhovich, M.Baroni, arXiv:1612.07182 (2016)
- [117] A.Chowdhury, J.R. Kubricht, A.Sood, P. Tu, A. Santamaria-Pang, in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)* (IEEE, 2020), pp. 1604–1607
- [118] A.Santamaria-Pang, J.Kubricht, A.Chowdhury, C. Bhushan, P. Tu, in *Intl. Conf. on Medical Image Comp. and Computer-Assisted Intervention* (Springer, 2020), pp. 326–334
- [119] I.Goodfellow, J.Pouget-Abadie, M.Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, in *Advances in Neural Information Processing Systems 27*, (Curran Associates, Inc., 2014), pp. 2672–2680.
- [120] X.Chen, Y.Duan, R.Houthoofd, John Schulman, I. Sutskever, P. Abbeel, in *Adv. in Neural Inf. Proc. Sys. 29*, pp. 2172–2180. arXiv:1606.03657
- [121] J.Yu, Z.Lin, J.Yang, X. Shen, X. Lu, T. S. Huang, in *2018 IEEE/CVF Conf. on Comp. Vis and Pat. Recog* (2018), pp. 5505–5514
- [122] J.Zhu, T.Park, P.Isola, A.A. Efros, in *2017 IEEE Intl. Conf. on Comp. Vis (ICCV)* (2017), pp. 2242–2251
- [123] T.Park, M.Liu, T.Wang, J. Zhu, in *2019 IEEE/CVF Conf. on Comp. Vis. and Pat. Recog. (CVPR)* (2019), pp. 2332–2341
- [124] P.Isola, J.Zhu, T.Zhou, A. A. Efros, in *2017 IEEE Conf. on Comp. Vis. and Pat. Recog. (CVPR)* (2017), pp. 5967–5976
- [125] Z.Yang, X.Li, L.Catherine-Brinson, A. N. Choudhary , W. Chen , A. Agrawal, *J. of Mech. Des.*, **140**(11) (2018)
- [126] X.Li, Z.Yang, L.C. Brinson, A. Choudhary , A. Agrawal , W. Chen, in *ASME 2018 Intl. Des. Eng. Tech. Conf. and Comp. and Info. in Eng. Conf.* (American Society of Mechanical Engineers Digital Collection, 2018)
- [127] L.Mosser, O.Dubrulle, M.J. Blunt, arXiv:1704.03225 (2017)
- [128] qubvel/segmentation_models: Segmentation models with pretrained backbones. Keras and TensorFlow Keras. https://github.com/qubvel/segmentation_models. Accessed 02 April 2021.
- [129] S. Akers, E. Kautz, A. Trevino-Gavito, M. Olszta, B. Matthews, L. Wang, Y. Du, S. Spurgeon, (2021) <https://doi.org/10.21203/rs.3.rs-346102/v1>
- [130] S. V. Kalinin, O. Dyck, A. Ghosh, Y. Liu, R. Proksch, B. G. Sumpter, M. Ziatdinov, arXiv:2010.09196 (2020)
- [131] X. Wang, J. Li, H. D. Ha, J. C. Dahl, J. C. Ondry, I. Moreno-Hernandez, T. Head-Gordon, and A. P. Alivisatos, *JACS Au*, **1** (2021) <https://doi.org/10.1021/jacsau.0c00030>
- [132] S. Madireddy, D.W. Chung, T. Loeffler, S.K. Sankaranarayanan, D.N. Seidman, P. Balaprakash, O. Heinonen, *Sci. Rep.* **9** (1) 1 (2019)
- [133] L. Mosser, O. Dubrulle, M.J. Blunt, arXiv:1704.03225
- [134] B.S.S. Pokuri, S. Ghosal, A. Kokate, S. Sarkar, and B. Ganapathysubramanian, *npj Comp. Mater.* **5** (95) (2019)
- [135] C. Yeung, J.M. Tsai, B. King, Y. Kawagoe, D. Ho, M. W. Knight, and A.P. Raman, *ACS Photonics*, **7** (2309-2318) (2020)
- [136] S. Liu, B. Kailkhura, J. Zhang, A. M. Hiszpanski, E. Robertson, D. Loveland, T. Yong-Jin Han, arXiv:2007.08631 (2020)
- [137] A. Belianinov, A. V. Ievlev, M. Lorenz, N. Borodinov, B. Doughty, S. V. Kalinin, F. M. Fernández, and O. S. Ovchinnikova, *ACS Nano*, **12** (12), (2018)
- [138] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee and A. Y. Ng, *Multimodal Deep Learning*. ICML (2011)